ECONOMETRICS RETAKE EXAM

SOLUTIONS PART B

Universidad Carlos III de Madrid

June 17, 2024

Write your name and group in each answer sheet. Answer the questions in 90'.

1. (40%) The scrap rate for a manufacturing firm is the number of defective items-products ($scrap$) that must be discarded, out of 100 produced. We are interested in using the scrap rate to measure the effect of worker training on productivity. A sample of firms is used to obtain the following OLS model estimate

$$\begin{aligned} \ln\widehat{(scrap_i)} &= \underset{(4.57)}{11.74} - \underset{(0.019)}{0.042} \cdot hrsemp_i - \underset{(0.370)}{0.951} \cdot \ln(sales_i) + \underset{(0.360)}{0.992} \cdot \ln(employ_i), \\ n &= 43,\ RSS = 65.91, \end{aligned}$$

where ln means Napierian logarithm, $hrsemp$ is the annual hours of training per employee, $sales$ is the annual firm sales (in dollars) and $employ$ is the number of firm employees. It is reported that $\widehat{Cov}\left(\hat{\beta}_{\ln(sales)}, \hat{\beta}_{\ln(employ)}\right) = -0.11$.

a. (1/5) (i) State the assumption you need on the error term of the model to interpret the coefficient of ln($sales$) as the expected elasticity of $scrap$ with respect to $sales$ (50%). (ii) Which additional assumption do we need in order to correctly interpret the estimated coefficient of $hrsemp$ as follows: "on average, $scrap$ decreases in a 4.2% when $hrsemp$ increases in one hour and the rest of explanatory variables remain fixed"? (50%).

SOLUTION: (i) We must assume that $\mathbb{E}\left(e^{u_i} \middle| hrsemp_i, sales_i, employ_i\right) = E\left(e^{u_i}\right) = A$, where u_i is the error term in the model. This is almost equivalent to assume that u_i is independent of all the explanatory variables. Notice that

$$\begin{aligned} \ln E\left(scrap \middle| hrsemp, sales, employ\right) &= \ln(A) + \beta_{hrsemp} hrsemp + \beta_{\ln(sales)} \ln(sales) \\ &\quad + \beta_{\ln(employ)} \ln(employ), \end{aligned}$$

where $A = E(e^{u_i})$. The expected elasticity of *scrap* with respect to *sales* is

$$\begin{aligned} \xi_{scrap,sales} &= \frac{\partial}{\partial \ln(sales)} \ln E(scrap|\, hrsemp, sales, employ) \\ &= \beta_{\ln(sales)}. \end{aligned}$$

(ii) We must also assume that the relative variation of $E(scrap|\, hrsemp, sales, employ)$, implied by increasing *hrsemp* in one hour, is small. More formally,

$$\begin{aligned} &\frac{E(scrap|\, hrsemp+1, sales, employ) - E(scrap|\, hrsemp, sales, employ)}{E(scrap|\, hrsemp, sales, employ)} \\ = &\frac{\Delta E(scrap|\, hrsemp, sales, employ)}{E(scrap|\, hrsemp, sales, employ)} \end{aligned}$$

must be small.

b. (2/5) Consider the following slight reformulation of the estimated model

$$\begin{aligned} \ln\widehat{(scrap_i)} &= \underset{(4.57)}{11.74} - \underset{(0.019)}{0.042} \cdot hrsemp_i - \underset{(0.370)}{0.951} \cdot \ln\left(\frac{sales_i}{employ_i}\right) + \underset{(??)}{0.041} \cdot \ln(employ_i), \\ n &= 43,\ RSS = 65.91, \end{aligned}$$

Obtain the omitted standard error of the OLS coefficient of $\ln(employ)$.

SOLUTION: Define

$$\begin{aligned} \hat{\theta} &= \hat{\beta}_{\ln(sales)} + \hat{\beta}_{\ln(employ)} \\ &= -0.951 + 0.992 \\ &= 0.041 \end{aligned}$$

Thus,

$$\begin{aligned} SE\left(\hat{\theta}\right) &= \sqrt{\widehat{Var}\left(\hat{\beta}_{\ln(sales)}\right) + \widehat{Var}\left(\hat{\beta}_{\ln(employ)}\right) + 2\widehat{Cov}\left(\hat{\beta}_{\ln(sales)}, \hat{\beta}_{\ln(employ)}\right)} \\ &= \sqrt{SE^2\left(\hat{\beta}_{\ln(sales)}\right) + SE^2\left(\hat{\beta}_{\ln(employ)}\right) + 2\widehat{Cov}\left(\hat{\beta}_{\ln(sales)}, \hat{\beta}_{\ln(employ)}\right)} \\ &= \sqrt{0.370^2 + 0.360^2 + 2 \cdot (-0.11)} \\ &= 0.215\,64. \end{aligned}$$

c. (2/5) Derive the 95% confidence interval for the expected *scrap* rate elasticity of $sales - to - employee$ ratio (20%). Using this interval, test the hypothesis that a 5% increase in $sales/employ$ is associated with an average 5% drop in the scrap rate (80%).

SOLUTION: Expected *scrap* rate elasticity of $sales - to - employee$ ratio,

$$\hat{\xi}_{scrap,sales/employ} = \hat{\beta}_{\ln(sales)} = -0.951.$$

Thus, the 95% confidence interval is

$$\begin{aligned} CI\left(\hat{\xi}_{scrap,sales/employ}\right) &= \hat{\beta}_{\ln(sales)} \pm 1.96 \cdot SE\left(\hat{\beta}_{\ln(sales)}\right) \\ &= -0.951 \pm 1.96 \cdot 0.370 \\ &= [-1.676\,2, -0.2258]. \end{aligned}$$

We are asked to test that the expected elasticity is -1, which can be stated as

$$H_0 : \beta_{\ln(sales)} = -1 \ \text{ vs } H_1 : \beta_{\ln(sales)} \neq -1.$$

Since the value -1 is inside the 95% confidence interval, we cannot reject H_0 at the 5% significance level. Note: Using the test, without stating the hypothesis and applying the confidence interval has only half grade.

2. (30%) Consider the multiple regression model,

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + U,$$

where $E(U|X) = 0$, and the rest of conditions for consistency and asymptotic normality of OLS estimators hold. We have *iid* observations $\{Y_i, X_i\}_{i=1}^n$ of (Y, X).

a. (1/3) Provide the value of X where the partial effect of X on Y changes sign (20%). Then, explain carefully how you can test, at the 5% of significance, that this value is bigger than 1 (80%).

SOLUTION: The value is

$$X^* = -\frac{\beta_1}{2 \cdot \beta_2},$$

which is estimated by

$$\hat{X}^* = -\frac{\hat{\beta}_1}{2 \cdot \hat{\beta}_2}.$$

The hypothesis to test is

$$\begin{aligned} H_0 &: X^* = 1 \text{ vs } H_1 : X^* > 1 \\ &\Updownarrow \\ H_0 &: -\frac{\beta_1}{2 \cdot \beta_2} = 1 \text{ vs } H_1 : -\frac{\beta_1}{2 \cdot \beta_2} > 1 \\ &\Updownarrow \\ H_0 &: \beta_1 + 2 \cdot \beta_2 = 0 \text{ vs } H_1 : \beta_1 + 2 \cdot \beta_2 < 0. \end{aligned}$$

This one-sided test is implemented using the $t - statistic$

$$t = \frac{\hat{\theta}}{SE\left(\hat{\theta}\right)},$$

where $\hat{\theta} = \hat{\beta}_1 + 2 \cdot \hat{\beta}_2$, and $\hat{\beta}_1$ and $\hat{\beta}_2$ are the OLS estimators of β_1 and β_2, respectively, and

$$\begin{aligned} SE\left(\hat{\theta}\right) &= \sqrt{\widehat{Var}\left(\hat{\beta}_1\right) + 2^2 \cdot \widehat{Var}\left(\hat{\beta}_1\right) + 2 \cdot 2 \cdot \widehat{Cov}\left(\hat{\beta}_1, \hat{\beta}_2\right)} \\ &= \sqrt{\widehat{Var}\left(\hat{\beta}_1\right) + 4 \cdot \widehat{Var}\left(\hat{\beta}_1\right) + 4 \cdot \widehat{Cov}\left(\hat{\beta}_1, \hat{\beta}_2\right)} \end{aligned}$$

The estimated variances and covariances must be robust to the presence of heteroskedasticity. So, we reject H_0 when $t < -Z_{0.05} = -1.645$.

Note: Full credit will be given if it is explained how to calculate the t statistic using the transformed model, but no additional credit will be given for doing that.

b. (1/3) Suppose that you incorrectly estimate the partial effect of X by the OLS slope estimator in the simple regression model, i.e.

$$\hat{\gamma}_1 = \frac{\widehat{Cov}\,(Y, X)}{\widehat{Var}\,(X)}.$$

Derive a formula for the asymptotic bias of $\hat{\gamma}_1$ as an estimate of β_1 (30%). Can the asymptotic bias be positive when $\beta_2 < 0$ and $X > 0$? (70%).

SOLUTION: Define the OLS residuals

$$\hat{U}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i - \hat{\beta}_2 X_i^2$$

Notice that

$$\begin{aligned} \hat{\gamma}_1 &= \frac{\widehat{Cov}\left(\hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_2 X^2 + \hat{U}, X\right)}{\widehat{Var}\,(X)} \\ &= \hat{\beta}_1 + \hat{\beta}_2 \frac{\widehat{Cov}\,(X^2, X)}{\widehat{Var}\,(X)} \\ &\to \beta_1 + \beta_2 \frac{Cov\,(X^2, X)}{Var\,(X)} \text{ w.p.1. as } n \to \infty \end{aligned}$$

Thus, the asymptotic bias is

$$Asympt.\ Bias = \beta_2 \frac{Cov\,(X^2, X)}{Var\,(X)} < 0$$

if $\beta_2 < 0$ and $X > 0$.

c. (1/3) Which assumption for the consistency of the OLS estimate $\hat{\beta}_1$ is not satisfied when X takes only the values 0 and 1? Justify carefully your answer.

SOLUTION: Since $X = X^2$ when X takes only the values 0 and 1, the model suffers of perfect multicollinearity.